Testing Concordance of Clinical Characteristics in Familial Studies with Application to Inflammatory Bowel Diseases

Jun Yan
Department of Statistics and Actuarial Science, University of Iowa
Iowa City, IA 52242, USA
jyan@stat.uiowa.edu
Cyrus P. Tamboli
Division of Gastroenterology, Department of Internal Medicine, University of Iowa
Iowa City, IA 52242, USA

July 5, 2007

Summary

The etiology of chronic Inflammatory Bowel Diseases (IBD) remains unknown, with both genetic and environmental risk factors having been implicated. A recent collaborative study of IBD provides clinical data from families with three or more affected first-degree relatives. The scientific question is whether specific clinical characteristics aggregate among affected individuals within families. Gastroenterological researchers have examined the number of concordant familial pairs in familial aggregation studies, but methods and results have been discrepant. This article investigates concepts of concordance and gives a comprehensive statistical treatment for testing concordance of various clinical traits in familial studies. For dichotomous traits, the distribution of this statistic under the null hypothesis of no familial aggregation is obtained by three methods: asymptotic, probability generating function, and permutation. The permutation method is extended to analyze aggregation for non-dichotomous traits and co-aggregations between two traits. We apply the permutation method to analyze the aforementioned multiply-affected IBD family data. Evidence is found for familial clustering of various traits, some of which are not revealed in existing studies. Such analyses provide a basis for investigating the dependence of trait aggregation upon genetic or environmental risk factors.
Keywords: concordance; probability generating function; familial association; inflammatory bowel disease; random permutation

1. INTRODUCTION

Crohn's disease (CD) and ulcerative colitis (UC) are two major types of chronic inflammatory bowel diseases (IBD). They may cause severe watery or bloody diarrhea, abdominal pain, malnutrition, and even death. Clinical differences exist between CD and UC but also among subtypes within each disease. The etiology of IBD remains poorly understood. It is likely that multiple risk factors and/or etiologies are operational (Russell and Satsangi, 2004). Further identification of homogeneous subtypes would be useful for epidemiological studies, particularly genetic association studies which depend upon homogeneous subgroups for identifying associations.

The International Organization for IBD (IOIBD) Multiplex Families Project is an international collaborative study of IBD families. The study has developed a database of families

comprising three or more affected first-degree relatives from 17 centers world-wide: Copenhagen, Dublin, Iowa City, Keio, Lille, Leuven New York, Orebro, Oslo, Oxford, Paris St. Antoine, Regensburg, Rome, San Giovanni Rotondo, Toronto, Vienna, and Winnipeg. All of the families are Caucasian except one from Japan. The underlying assumption of such a data collection scheme is that genetic and environmental risk factors for IBD are stronger among multiply-affected families. This may also manifest itself as shared disease characteristics among affected relatives. Medical records of all affected individuals were carefully reviewed to verify diagnosis. Ascertainment of subjects was not based upon any of the specific clinical characteristics to be analyzed here, except for having the diagnosis of IBD. We emphasize that only IBD-affected subjects are included and that all families in the dataset have three or more affected first-degree relatives. There are 238 families and 769 subjects, among which 197, 32, 5, 3, and 1 families have, respectively, 3, 4, 5, 6, and 7 affected members. Each family represents an independent pedigree; no extended pedigrees were included as much as could be determined. Researchers are interested in testing the familial aggregation of IBD clinical characteristics, which may lead to better understanding of the etiology of IBD. The clinical traits of interest are listed in Table 1; see the Application section for more details.

Gastroenterologists have employed concordance rates to study familial aggregation in this context. Previous concordance studies have led to various conclusions (Bayless et al., 1996; Lee and Lennard-Jones, 1996; Peeters et al., 1996). The definition of concordance has sometimes differed. Some (Lee and Lennard-Jones, 1996) have defined concordance as unanimous trait sharing among all family members. Others (Bayless et al., 1996; Peeters et al., 1996) defined the proportion of concordant familial pairs among all familial pairs. Under the untestable assumption of independence between families, which is assumed in this study as in many other familial studies, concordance analysis of the first type, at the "family level", is easier than the second type, at the "relative pair" level. This article addresses concordance analysis at the level of relative pairs. The association of specific disease traits among affected relative pairs within families is of explicit research interest.

We illustrate the set-up of the statistical problem using a dichotomous trait. Let Z be the indicator variable for a certain clinical trait: $Z = 1$ for presence and $Z = 0$ for absence of the trait. A relative pair is concordant in Z if both individuals have this clinical trait ($Z = 1$). Absence of a trait in both members is not considered to represent any shared risk factor under this definition and so is not considered as concordant. That is, only a matching 1–1 pair is treated as concordance; neither 1–0 nor 0–0 is treated as concordance. Let $p = \Pr(Z = 1)$ be the prevalence of this clinical trait among all affected subjects in the study. For simplicity, assume that all families are of size k and that we observe this variable for n_k families. Among all $n_k\binom{k}{2}$ within-family pairs, we count the number of concordant relative pairs X_{n_k}. The statistical question is whether the observed X_{n_k} is significantly greater than expected under the null hypothesis:

$$H_0 : \text{the presence of the clinical trait } Z \text{ is randomly assorted across families.} \qquad (1.1)$$

Published concordance analyses of IBD in the gastroenterology literature have mostly used population-based estimates of phenotypic trait prevalence to construct reference distributions for concordance rates under H_0 in (1.1). Subjects with diagnosis of IBD in a multiplex family study however, are clearly a high-risk group, which makes it inappropriate

to use population-based reference distributions for these clinical traits. Furthermore, the reference distributions have been mostly based on asymptotic results, whose validity may be in question for rare traits with p close to zero. Practitioners need guidelines on when it would be appropriate to apply the asymptotic results. If the asymptotic results are inappropriate, a robust method which is not sensitive to rare traits is needed.

In the IOIBD study, there are other practical issues to be solved. First, there are non-dichotomous traits such as multichotomous, ordinal, or continuous traits. For example, the disease "behavior" type in Crohn's disease (disease "behavior" is a commonly utilized classification scheme in CD to describe key clinical features of the disease) is a multichotomous variable with 4 possible outcomes, the number of surgeries is an ordinal variable, and age at diagnosis is a continuous variable. Second, some groups of relative pairs are of particular interest. For example, siblings are the first-degree relatives at highest risk of developing IBD (Russell and Satsangi, 2004). Third, it is clinically significant to determine if concordance in Z_1 is associated with concordance in Z_2. This need arises when two clinical characteristics are both determined to aggregate within families and the two traits are also known to be clinically associated with each other (Louis et al., 2003). Demonstration of this implies co-aggregation of the two clinical traits (Hudson et al., 2001).

The goals of this article are three-fold. First, for dichotomous traits, we provide guidelines on using the normal approximation of the distribution of X_{n_k} under H_0 by investigating three methods: asymptotic, probability generating function (PGF), and permutation. Second, we extend the permutation method to solve the practical issues in the concordance analyses of the IOIBD study. Finally, we analyze the multiplex data from the IOIBD study using the proposed permutation method.

This article is organized as follows. Dichotomous traits are considered in Section 2: the large sample result is presented in Section 2.1; the PGF method is presented in Section 2.2; and the random permutation method is presented in Section 2.3. Extensions of the permutation method in three directions are presented in Section 3: non-dichotomous traits, partial permutation, and bivariate permutation for concordance association. These methods are applied to IOIBD data in Section 4. A discussion concludes in Section 5.

2. TESTING FOR DICHOTOMOUS TRAITS

In this section, we consider testing H_0 in (1.1) for a dichotomous trait Z using the asymptotic method, PGF method, and permutation method. For simplicity, assume that there are n_k families of size k and let X_{n_k} be the number of concordant familial pairs. The null distribution of X_{n_k} obtained from the asymptotic method and the PGF method are unconditional, depending on a consistent estimate of the clinical trait prevalence p. In contrast, the null distribution obtained from the permutation method is conditional given the number of individuals with $Z = 1$.

2.1 Asymptotic method

For a family of size $k \geq 2$, consider N_k, the number of concordant pairs within this family. Let $\mu_k = E(N_k)$ and $\sigma_k^2 = \mathrm{Var}(N_k)$. It can be shown that $\mu_k = \binom{k}{2}p^2$ and

$$\sigma_k^2 = \begin{cases} p^2(1-p^2) & k = 2, \\ \frac{1}{4}k(k-1)p^2[2+(k-2)p\{4+(k-3)p\}] - \binom{k}{2}^2 p^4 & k \geq 3. \end{cases} \tag{2.1}$$

A proof is sketched in Appendix A. An alternative but equivalent form of (2.1) is given in Hunt et al. (1986). Let $\{N_{k,i} : i = 1, \ldots, n_k\}$ be a sequence of independent random variables with the same probability distribution as N_k. The total number of within-family concordance pairs is then $X_{n_k} = \sum_{i=1}^{n_k} N_{k_i}$. By the central limit theorem, when n_k is large, X_{n_k} is approximately

$$X_{n_k} \sim N(n_k\mu_k, n_k\sigma_k^2). \tag{2.2}$$

In a more practical setting, we observe n families, among which there are n_k families of size k for $k = 2, \ldots, K$. Suppose the proportion of families with size k, n_k/n, goes to some limit π_k for large n. Then, as n gets large, the total number of within-family concordance pairs from all n families, $X_n = \sum_{k=1}^{K} X_{n_k}$, is approximately

$$X_n \sim N\left(\sum_{k=1}^{K} n\pi_k\mu_k, \sum_{k=1}^{K} n\pi_k\sigma_k^2\right). \tag{2.3}$$

Application of (2.3) needs guidelines for when the approximation is adequate. A preliminary analysis of the IOIBD study data (Tamboli et al., 2005) used a jackknife method to obtain the variance estimate of X_n. The inference, however, is still based upon a normal approximation, which may be inherently inappropriate if the normal approximation is poor. This is similar to the case of normal approximation of a binomial distribution, where it is well-known that the mean of the binomial distribution needs to be at least 5 for the normal approximation to work well.

In order for (2.3) to be a good approximation, the normal approximation of X_{n_k} in (2.2) needs to be good for each k. Assume that n_kp^2 goes to some constant as $n \to \infty$ and $p \to 0$. Heuristically, normal approximation of the distribution of a positive discrete random variable can only work well if zero is at least two standard deviations away from the mean. That is, we require $n_k\mu_k > 2\sqrt{n_k\sigma_k^2}$. Note that as $p \to 0$, the first term of σ_k^2 in (2.1) is of order p^2 while other terms are of order $o(p^2)$. Therefore, a guideline for small p can be obtained as

$$n_kp^2 > 4\binom{k}{2}^{-1}. \tag{2.4}$$

Therefore, we need n_kp^2 to be at least 5, 2, and 1, respectively, for family of size $k = 2$, 3, and 4. Continuity correction is critically important when n_kp^2 is small. In cases where n_kp^2 is extremely small, the exact distribution of X_{n_k} is very skewed and a normal approximation is hopeless. This guideline (2.4) can be graphically illustrated by comparing the normal approximation to the exact distribution.

Figure 1 presents the normal approximation overlapped with the exact distribution of X_{n_k} obtained from the PGF method for families of size $k = 3$. The number of families n_k

takes two values: 100 and 200. The prevalence p takes three values 0.1, 0.2, and 0.4. From Figure 1, one observes that the approximation improves as p increases or n_k (and hence n) increases. For all three prevalence values, the normal approximation density is higher at the left tail and lower at the right tail than the exact distribution. Consequently, the p-value of an observed value of X_{n_k} on the right tail from an normal approximation is lower than the true p-value, which, particularly for small n and p, may lead to false discoveries of significance.

In practice, parameter p is unknown, and needs to be substituted with a consistent estimator $\hat{p}$. A convenient choice for $\hat{p}$ is the usual proportion estimator for binomial distributions.

2.2 Probability Generating Function Method

When the prevalence p is known, the exact distribution of X_{n_k} and X_n can be obtained from the PGF method. A similar approach has been developed in a different context to test the clustering of affected siblings and has been applied to IBD by Hugot et al. (2002).

For a discrete random variable N taking values in $\{0, 1, 2, \ldots\}$, the PGF G of N is defined by

$$G_N(t) = E(t^N) = \sum_{n=0} \Pr(N = n) t^n.$$

Consider N_k, the number of concordant pairs within a family of size k. The PGF of N_k is

$$G_{N_k}(t) = \left[\binom{k}{0} (1-p)^k + \binom{k}{1} p(1-p)^{k-1} \right] t^0 + \sum_{j=2}^{k} \binom{k}{j} p^j (1-p)^{k-j} t^{\binom{j}{2}}.$$

Then by independence between families, the PGF of $X_{n_k} = \sum_{i=1}^{n_k} N_{k,i}$ is

$$G_{X_{n_k}}(t) = [G_{N_k}(t)]^{n_k}. \tag{2.5}$$

This function is a polynomial of t. The probability mass function of X_{n_k} can be determined from (2.5):

$$\Pr(X_{n_k} = x) = \text{the coefficient of } t^x \text{ in } G_{X_{n_k}}(t).$$

Similarly, the PGF of $X_n = \sum_{k=1}^{K} X_{n_k}$ is

$$G_{X_n}(t) = \prod_{k=1}^{K} [G_{N_k}(t)]^{n_k},$$

and the exact distribution of X_n is determined by

$$\Pr(X_n = x) = \text{the coefficient of } t^x \text{ in } G_{X_n}(t).$$

The PGF method can be implemented in any software that can perform symbolic calculations, such as `Mathematica` (Wolfram Research, Inc., 2005). For illustration, the exact distributions of X_{n_k} are presented in Figure 1 in contrast to the normal approximations for families of size $k = 3$. A sample `Mathematica` session is given in Appendix B to show its ease of use in practice. Similar to the asymptotic method, this method depends on the unknown parameter p, which needs to be substituted with a consistent estimator $\hat{p}$.

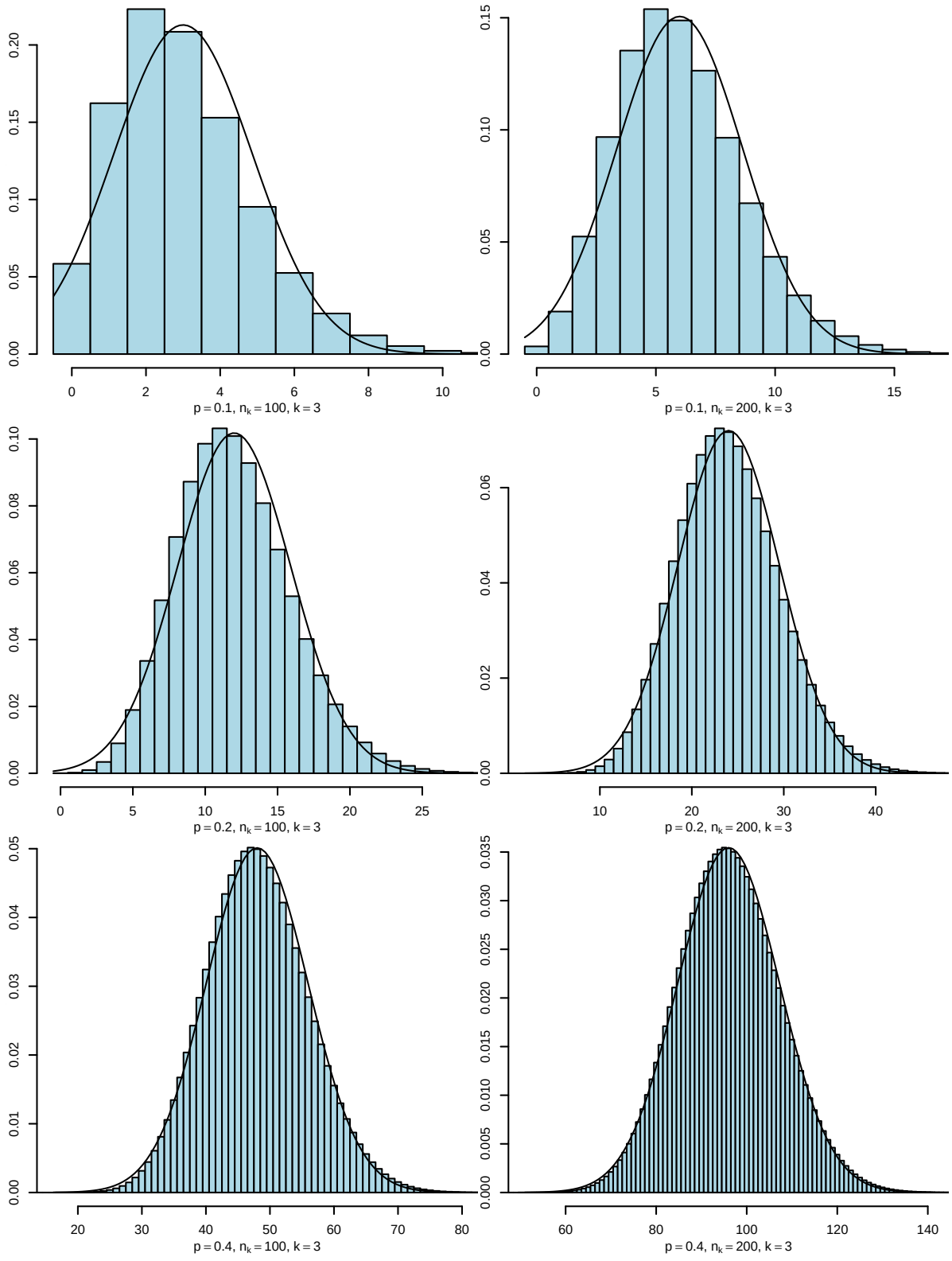

Figure 1: Normal approximation and exact distribution of the number of concordant relative pairs in families of size $k = 3$. The number of families n_k is 100 and 200. The clinical characteristic prevalence p is 0.1, 0.2, and 0.4.

2.3 Permutation Method

Resampling methods are an important tool in medical research (Berger, 2000; Good, 2001; Pesarin, 2001). An intuitive perspective of the permutation method is to view the phenotypic trait Z of a subject as a label. Under the null hypothesis H_0 of no familial clustering, the labels of individual subjects are exchangeable. That is, the label Z can be randomly assigned across families. Each permutation of the label Z is a way for such assignment. For each permutation, the test statistic X_n can be computed. Note that the number of presence labels ($Z = 1$) remains unchanged from permutation to permutation. Therefore, the null distribution of X_n obtained by exhausting all $(\sum_{i=1}^{K} n_k k)!$ permutations is conditional, given the total number of presences. This distribution is used to evaluate the significance of the observed statistic $X_n = x_n$ and give its p-value. When families have different sizes, the permutation method can be applied without any change to give the significance of the observed statistic X_n.

For large n, however, it is impossible to exhaust all the permutations. In practice, we draw a large number of random samples from all possible permutations to approximate the conditional null distribution of X_n. Monte Carlo error is introduced in this procedure and can be controlled by increasing the number of random permutations B. There are methods to determine the necessary size B for decision making at a given significance level α (see, for example, Besag and Clifford, 1991; Nettleton and Doerge, 2000). In this article, we use $B = 10,000$, which gives stable results in the analyses and leaves rooms for Bonferroni adjustment for multiple tests.

Under the same configurations as in Figure 1, Figure 2 presents a Monte Carlo approximation of the conditional null distribution of X_{n_k} using $B = 10,000$ random permutations, given that there are $n_k kp$ individuals with $Z = 1$. The unconditional asymptotic normal approximation in Figure 1 is overlapped to highlight the difference between the unconditional and conditional distribution. The approximate conditional distributions are clearly tighter than the unconditional distributions. This is not surprising, since the number of presences is fixed at $n_k kp$ in all permutations. The skewness properties of the distributions in Figure 2 are similar to those in Figure 1. The asymmetry is alleviated as p increases or n increases. For $p \geq 0.2$ and $n \geq 100$, the distributions look more symmetric.

The permutation method differs from the asymptotic method and PGF method by not requiring estimation of p. Therefore, it is not affected by the sample size n and the phenotypic trait prevalence p. The method is easy to implement in any computing environment. It does require a large number of permutations for a good approximation of the reference distribution, but for the problem considered in this article, it is not computationally demanding by modern statistical computing standards.

An additional advantage of the permutation is that the method is still valid when there are missing values (NA), which may or may not be missing completely at random. For instance, missing may be more likely to happen when a trait is presence for a subject than otherwise. The permutation method can treat NA as a special value which does not contribute to concordance. Under H_0, subjects are still exchangeable, whether or not their clinical traits are missing. The null distribution can still be approximated by random permutations to assess the significance of the observed number of concordances.

It is a common need to control the effect of some demographic factors such gender and

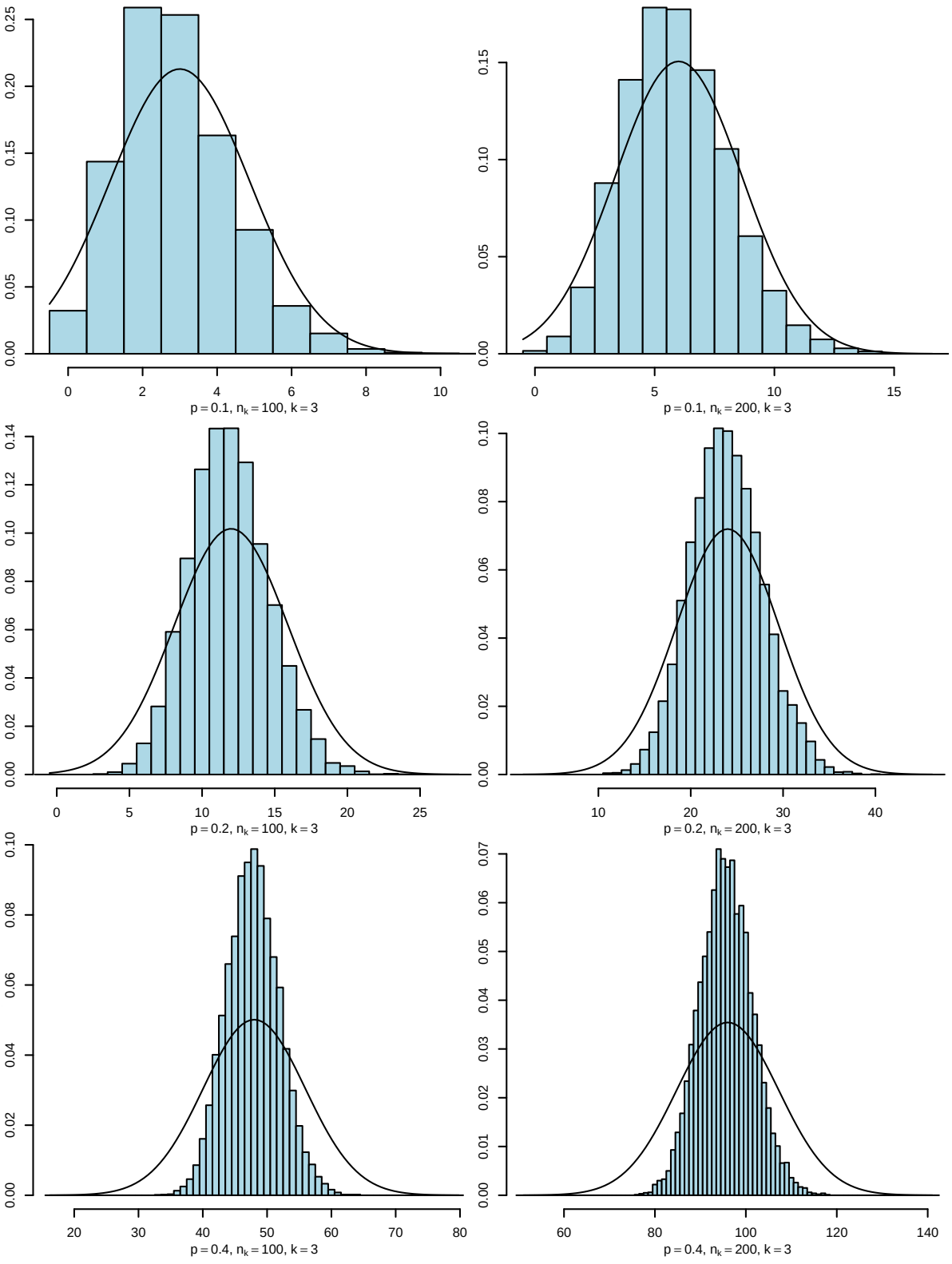

Figure 2: Approximation of the conditional distribution of the number of concordant relative pairs in families of size $k = 3$. The size of the random permutation sample is $B = 10,000$. The number of families n_k is 100 and 200. The clinical characteristic prevalence p is 0.1, 0.2, and 0.4. The unconditional asymptotic normal approximation in Figure 1 is overlapped to highlight the difference between the conditional and unconditional distributions.

age. For categorical covariates with a small number of levels such as gender, one simply constrains the permutation to subjects who are in the same level of the covariate (Lynch et al., 1981, 1986; Schwartz et al., 1988). For a continuous covariate, one need to stratify it into several categories and constrain the permutation of a clinical trait within each category. For example, the variable age can be discretized into young and old groups by the medium age. To otherwise incorporate covariates, regression analysis is generally required.

3. EXTENDING THE PERMUTATION METHOD

3.1 Concordance of Multichotomous, Ordinal and Continuous Traits

In the IOIBD study, some traits are multichotomous, continuous, or ordinal. Let Z now be such a trait. The more general null hypothesis is

$$H_0 : \text{the clinical character } Z \text{ is randomly assorted across families.} \tag{3.1}$$

The testing statistic is still the number of concordant relative pairs X_n, given an appropriate definition of concordant pair in this trait. A concordant pair in a multichotomous trait, disease "behavior" type, may be defined as a pair who have the same subtype of the trait. A concordant pair in a continuous trait, age of diagnosis, may be defined as a pair whose age at diagnosis are at most 2 years apart. A concordant pair in an ordinal trait, number of surgeries, may be defined as a pair who both have had surgery and whose number of surgeries differ by at most 1. With these new concordance definitions, one can proceed with the permutation method to test H_0 in (3.1).

It is also possible to use some other statistic which measures the sum of pairwise distances between family members. Let Z_{ij} be a continuous trait of the jth member in family i. An example distance measure is $S_n = \sum_{i=1}^{n} \sum_{j<k}^{m_i} (Z_{ij} - Z_{ik})^2$, where m_i is the size of family i.

3.2 Bivariate Permutation for Concordance Association

The permutation method can be extended to investigate the association of concordance in two clinical characteristics. It may be hypothesized that if concordance in trait Z_1 and concordance in trait Z_2 are seen concurrently more often than expected by chance, then the two traits co-aggregate. The null hypothesis is

$$H_0 : \text{the familial aggregation of } Z_1 \text{ and } Z_2 \text{ are independent.} \tag{3.2}$$

For ease of presentation, consider first the case where there are only n_k families of the same size k. A reasonable test statistic is the number of relative pairs Y_{n_k} that are simultaneously concordant in both Z_1 and Z_2. When H_0 in (3.2) is not true because of co-aggregation of these two characteristics, Y_{n_k} is likely to be higher than expected by chance. It is crucial to determine the reference distribution of Y_{n_k} under (3.2) without changing the familial association property of both variables as regards the concordances. A two-stage random permutation is thus designed for this purpose. In the first stage, permutation is done at the family level, i.e., the Z_1 labels of each family as a group are randomly re-assigned to a different family identity. In the second stage, the Z_1 labels are randomly permuted among individuals within each family. The random permutation of the Z_2 labels can be done in

the same way Clearly, this method preserves the number of concordant pairs in Z_1 and Z_2, but the number of pairs simultaneously concordant in Z_1 and Z_2 changes from permutation to permutation. A large number of such permutations provide a reference distribution for Y_{n_k} under the null hypothesis of no concordance association, given the observed number of concordant relative pairs in each trait. The observed number of simultaneous concordant pairs y_{n_k} is then positioned against this reference distribution to give a p-value.

In the general setting where the n families consist of n_k families of size k for $k = 2, \ldots, K$, we use the statistic Y_n, the total number of simultaneous concordant within-family pairs in Z_1 and Z_2. The reference distribution of Y_n is still obtained from the two-stage permutation, except that in the first stage, families of the same sizes are permuted separately. The method is valid when there are a few very large pedigrees. For example, in the IOIBD data, there is only 1 family of size 7 and variable values of subjects in this family are permuted within the family.

4. ANALYSIS OF THE IOIBD DATA

The clinical traits of interest in the IOIBD study are summarized in Table 1. There are 21 variables: 14 dichotomous, 4 multichotomous, 1 ordinal, and 2 continuous. The 4 multichotomous variables all have 3 or more states. These variables have different domains of relevance. For IBD and its two main types, CD and UC, a mark "$\surd$" indicates that a variable is relevant.

Table 2 presents frequency summaries for several aspects of the IOIBD data, such as the composition of gender, relation, and some multichotomous variables. Due to missing values, the total number of observations for each variable is not necessarily equal to the total number of subjects in the relevant domain.

Out of the 39 clinical traits to be analyzed for three relevant domains in Table 1, 25 have less than 5% missing values and only 3 have greater than 10% missing values. The 3 traits with heaviest missingness are all for CD: uppertract (23.7%), aza6mp (20.75%), and behaviour (16.8%). For traits with non-negligible amount of missing values, it is necessary to investigate whether the missingness was missing completely at random (MCAR) (Little and Rubin, 2002). For example, for upper tract involvement (uppertract), missing is more likely to happen when the trait is absent than when it is present. Fortunately, with missing value treated as a special value which does not contribute to concordance counts, the permutation method is still valid, although the power of detecting familial aggregation may be low.

To apply the permutation method on non-dichotomous traits, concordance in these variables needs to be defined. A relative pair is defined as concordant in a multichotomous variable if their values are both non-zero and match exactly. Note that variables patype and and smoker have state zero, indicating no involvement and non-smoker, respectively. A relative pair is defined as concordant in surgcount if their values of surgcount are both non-zero and have difference within a threshold. The two continuous variables, symptom duration (sympdur) and age at diagnosis (ageatdx) are constructed from patients' age, diagnosis time, and calendar time. A relative pair is defined as concordant in a continuous variable if the values of the variables in the pair have difference within a threshold. In this paper, we define the threshold for surgcount and other time-related continuous variables as 2 surgeries and 2 years, respectively.

Table 1: Summary of clinical trait types and their domains of relevance in the IOIBD study.

Variable Name	Variable Interpretation	IBD	CD	UC
	Dichotomous variables			
uppertract	upper tract involvement		√	
termileum	terminal ileum involvement		√	
rectum	rectal involvement			√
leftcolon	left colonic involvement			√
extcolon	extensive colonic involvement		√	√
perianal	perianal involvement		√	
ctrlarth	central arthritis	√	√	√
prphlarth	peripheral arthritis	√	√	√
cutaneous	cutaneous involvement	√	√	√
ocular	ocular involvement	√	√	√
surgery	history of resection surgery	√	√	√
aza6mp	azathioprine or 6 mercaptopurine use		√	
mtx	methotrexate use		√	
	Multichotomous variables			
dx	diagnosis (CD, UC, or indeterminate colitis)	√		
patype	type of perianal involvement (none, abscess, or fistula)		√	
behavior	CD "behavior" (stenosing, fistulizing, fistulizing & perianal, or inflammatory)		√	
smoker	smoking status (non, ex, or current)	√	√	√
	Ordinal variables			
surgcount	number of surgeries	√	√	√
	Continuous variables			
sympdur	symptom duration	√	√	√
ageatdx	age at diagnosis	√	√	√

Table 2: Frequency summaries of the IOIBD data.

Variable	Subtype	Frequency	Variable	Subtype	Frequency
diagnosis	CD	506	CD smoker	non	222
	UC	244		ex	113
	indeterminate	18		current	144
gender	female	404	UC smoker	non	129
	male	365		ex	73
relation	proband	238		current	30
	parents	97	CD behavior	stenosing	118
	siblings	316		fistulizing	42
	children	118		fistulizing (perianal)	35
smoker	non	361		inflammatory	226
	ex	188	CD patype	none	403
	current	180		fistula	40
				abscess	21

We first investigate the familial aggregation in each clinical trait. In order to have some control of the effect of some basic demographic factors, we categorize subjects into categories formed by the interaction of factors that we want to control for. The permutation is done with the constraint that only subjects in the same category are permuted. For the IOIBD data, out categorization is done from the interaction of three factors: 1) gender (2 levels, male and female), 2) age (7 levels, 20 or below, 20–30, 30–40, 40–50, 50–60, 60–70, and above 70), and 3) disease duration (6 levels, 0–5, 5–10, 10–15, 15–20, 20–25 and above 25). Inclusion of disease duration as a control variable is desired by the clinicians to guard against false determination of genetic anticipation (Picco et al., 2001). Since there are 20 missing observation for this grouping variable, the permutation tests are done with the rest 749 subjects. They are from 238 families and there are 849 relative pairs in total.

Tables 3 summarize the permutation test results for all three domains of IBD, CD, and UC. The hypothesis of no association is tested for three types of relatives pairs: all relative pairs, parent-child pairs, and sib pairs. For each test, we report the observed number of concordant relative pairs (NC), the mean of the statistic from $B = 10,000$ random permutation, and the p-value of the observed statistic under the null hypothesis of no familial aggregation. The reported p-values are one-sided and represent the probability of the number of concordant relative pairs being at least as many as the observed number NC. The observed number of concordant relative pairs in the tables can be as small as 1 or 0. The permutation method is not affected by the rare traits and remains valid, though the test result may not be significant.

Results of these permutations tests suggest which of the clinical characteristics studied aggregate within families. It is apparent that concordance is more prevalent among various CD traits than among UC traits, and among sibling pairs. These findings are consistent

Table 3: Summary of IBD familial aggregation testing with the permutation method. NC is the observed number of concordant relative pairs. Mean and P-value are the mean and one-sided p-value computed from 10,000 random permutations under no familial aggregation.

	Trait	All relative pairs			Parent-child pairs			Sib pairs		
		NC	Mean	P-value	NC	Mean	P-value	NC	Mean	P-value
IBD	dx	662	463.9	0.0001	149	98.2	0.0001	387	276.4	0.0001
	ageatdx	120	98.4	0.0076	10	10.9	0.6756	94	74.5	0.0074
	sympdur	113	111.4	0.4419	21	18.7	0.2993	76	78.5	0.6581
	smoker	387	293.1	0.0001	92	61.1	0.0002	235	176.7	0.0001
	ctrlarth	8	2.9	0.0120	0	0.6	1.0000	8	1.8	0.0019
	prphlarth	71	37.8	0.0001	15	8.5	0.0192	45	22.1	0.0001
	cutaneous	8	6.3	0.2787	0	0.8	1.0000	5	4.1	0.3870
	ocular	5	1.8	0.0315	1	0.3	0.2683	3	1.1	0.0967
	surgery	186	157.3	0.0009	32	35.3	0.7635	122	95.4	0.0010
	surgcount	78	68.5	0.0932	16	15.3	0.4567	48	40.4	0.1010
CD	ageatdx	73	61.5	0.0513	4	5.0	0.7530	63	50.7	0.0255
	sympdur	64	70.8	0.8607	7	7.7	0.6944	47	53.4	0.8916
	smoker	213	159.4	0.0001	46	28.7	0.0003	138	103.6	0.0001
	ctrlarth	4	2.0	0.1223	0	0.4	1.0000	4	1.3	0.0340
	prphlarth	44	23.8	0.0002	9	5.7	0.0917	29	14.2	0.0001
	cutaneous	8	5.1	0.1167	0	0.8	1.0000	5	3.3	0.2197
	ocular	1	1.1	0.6744	0	0.1	1.0000	1	0.7	0.5291
	surgery	154	139.4	0.0333	26	30.0	0.8564	100	86.7	0.0392
	surgcount	61	54.9	0.1816	12	11.8	0.5239	39	33.5	0.1555
	uppertract	15	7.0	0.0037	4	0.6	0.0047	10	5.1	0.0235
	termileum	289	255.8	0.0001	50	49.2	0.4773	187	161.1	0.0004
	perianal	22	8.8	0.0005	3	1.7	0.2379	17	5.7	0.0009
	patype	12	3.7	0.0012	3	0.9	0.0608	7	2.2	0.0118
	behaviour	190	125.4	0.0001	35	22.2	0.0010	117	78.7	0.0001
	aza6mp	75	61.0	0.0119	9	6.2	0.1815	54	44.1	0.0387
	mtx	4	1.8	0.0866	2	0.2	0.0138	2	1.2	0.3522
	extcolon	135	108.2	0.0002	17	16.9	0.5293	94	71.9	0.0007
UC	ageatdx	28	18.7	0.0097	5	3.5	0.2589	17	11.4	0.0359
	sympdur	24	21.3	0.2693	9	5.8	0.0842	13	13.0	0.5625
	smoker	93	78.4	0.0130	24	21.3	0.2646	50	41.9	0.0399
	ctrlarth	1	0.1	0.1343	0	0.0	1.0000	1	0.1	0.1011
	prphlarth	14	7.9	0.0174	5	2.5	0.0798	5	4.3	0.4308
	cutaneous	0	0.0	1.0000	0	0.0	1.0000	0	0.0	1.0000
	ocular	1	0.3	0.2801	0	0.1	1.0000	0	0.1	1.0000
	surgery	12	9.7	0.2297	1	2.4	0.9400	9	5.2	0.0692
	surgcount	7	5.2	0.2358	1	1.3	0.7566	4	3.2	0.3835
	rectum	0	0.0	1.0000	0	0.0	1.0000	0	0.0	1.0000
	leftcolon	48	44.8	0.2459	16	11.7	0.0880	28	24.1	0.1648
	extcolon	21	16.4	0.0867	1	3.6	0.9790	14	8.9	0.0380

with IBD literature which shows stronger familial and genetic influences for CD than for UC (Russell and Satsangi, 2004). Our methodology has also uncovered new familial associations not previously described. We report upon aggregation of many extra-intestinal features of IBD not previously reported, and we demonstrate interesting familial aggregations of some of these traits with respect to generational subsets. For example, peripheral arthritis (prphlarth) aggregation occur in sibling pairs but not parent-child pairs in CD group. Results such as these can deduce the higher importance of environmental influences on this trait since parent-child pairs share similar percentage of genes as sib-sib pairs (unless they are monozygotic twins, but very few such pairs are included in this cohort). As well, siblings are likely to share environments to a much greater extent than parent-child pairs. Firstly, parents have a unique exposure history even prior to their children's births. Second, sibs' school environments are likely to be similar if they are of similar ages, which is a unique environment from the parent's workplace.

In the first phase of this IOIBD project, the clinicians were most interested in exploring familial aggregation of all clinical traits rather than controlling the false discovery rate. Thus, multiple testing was considered a secondary issue in this analysis. Nevertheless, we have run the analysis with $B = 10,000$ random permutations such that the minimum possible p-value is 0.0001 to allow room for Bonferroni adjustment. There are 117 p-values reported in Table 3. One can make the Bonferroni adjustment by simply inflate each p-value by a scale of 117. This implies that only those p-values which are 0.0004 or below in Table 3 lead to a significance level of 0.05. Of note is that Bonferroni adjustment is by nature conservative and should not stop clinicians from further investigating those detected aggregations which become insignificant after the adjustment.

We now investigate the association of concordances in two clinical characteristics. This part of the analysis is for exploratory purpose. We use the two-stage permutation without grouping constraint. The permutation is done on all 769 subjects. Table 4 summarizes the testing results for IBD, CD, and UC traits. For each pair of traits, we report the observed number of concordant relative pairs in trait 1 (NC1) and trait 2 (NC2), and the observed number of relative pairs that are simultaneously concordant in both traits (NSC). Under the null hypothesis of no concordance association between traits 1 and 2, the mean of the statistic NSC and one-sided p-value of the observed NSC are obtained from the $B = 10,000$ random permutations. Only pairs of traits with p-values 0.10 or lower are shown.

It can be concluded for example, that concordance in the number of surgical resections (surgcount) significantly co-aggregates with concordance for the same diagnosis (dx) among family members in the entire IBD group, and co-aggregates with concordance for terminal ileum involvement (termileum) in the CD subset. These findings support well-recognized phenomena that CD individuals are at risk for multiple surgeries, while those with UC typically have no more than one intestinal resection, and that terminal ileum involvement in CD presents a higher risk for requiring multiple resections. In this example, the bivariate permutation analysis demonstrates significant familial co-aggregation for a continuous variable surgcount with two other key traits, despite its own lack of familial aggregation, at significance level 0.05, in univariate concordance analysis for IBD, CD or UC subgroups. Identification of such trait co-aggregations may have utility in testing genotype-phenotype correlations where phenotypic heterogeneity can often mask true associations, or as an alternative method for testing gene-environment interactions. Co-aggregation of two traits

Table 4: Summary of testing the association of concordance in two clinical traits. Only pairs with p-value below 0.10 are listed. NC1 and NC2 are the number of concordant relative pairs for the two traits, respectively. NSC is the number of relatives that are simultaneously concordant in the two traits. Mean and P-value are the mean and one-sided p-value computed from 10,000 random permutations under no association of concordance.

	Trait 1	Trait 2	NC1	NC2	NSC	Mean	P-value
IBD	dx	smoker	695	406	322	312.34	0.0956
	dx	ageatdx	695	120	102	93.28	0.0264
	ageatdx	dx	120	695	102	93.36	0.0304
	prphlarth	extcolon	75	210	24	17.18	0.0854
	cutaneous	extcolon	10	210	6	2.26	0.0237
	surgery	dx	196	695	172	153.56	0.0016
	surgery	ageatdx	196	120	36	26.83	0.0243
	surgery	extcolon	196	210	66	48.59	0.0073
	surgcount	dx	80	695	70	62.41	0.0259
CD	uppertract	ageatdx	15	73	5	2.26	0.0656
	uppertract	behav	15	192	11	5.46	0.0090
	termileum	behav	296	192	130	117.08	0.0158
	extcolon	behav	140	192	69	55.53	0.0118
	perianal	extcolon	22	140	11	6.88	0.0687
	patype	behav	12	192	12	5.65	0.0022
	patype	extcolon	12	140	8	3.48	0.0115
	prphlarth	uppertract	48	15	4	1.36	0.0595
	cutaneous	extcolon	10	140	6	2.86	0.0519
	surgery	ageatdx	160	73	29	23.32	0.0917
	surgery	extcolon	160	140	55	46.24	0.0714
	surgery	termileum	160	296	121	99.59	0.0001
	surgcount	termileum	63	296	48	39.71	0.0175
	aza6mp	smoker	78	225	42	34.79	0.0613
	aza6mp	extcolon	78	140	30	21.40	0.0289
	aza6mp	termileum	78	296	52	45.43	0.0965
UC	extcolon	smoker	21	95	15	9.96	0.0208
	surgery	extcolon	12	21	4	1.25	0.0495

might also be of clinical utility from a predictive standpoint. For example, an individual who is concordant with a sibling for extensive colonic involvement (extcolon) appears at higher risk of requiring surgery if the sib had surgical resection. Although the directionality of association between extensive colonic involvement and surgical resection seems clinically intuitive in this example and is suggestive that there is a causal relationship (i.e. extensive colonic involvement would predate and predict a higher likelihood of requiring surgery and also that lesser extent of involvement — leftcolon and rectum — were not associated with surgery in this bivariate analysis), this type of conclusion is not as straightforward for other sets of covariates. Causality, in general, cannot be inferred from this type of analysis. A formal causal inference along the line of Holland (1986) in the context of familial studies is worth investigating. Application in clinical and epidemiological studies are, for example, Little and Rubin (2000) and Maldonado and Greenland (2002). A recent book treatment is Rubin (2006).

5. DISCUSSION

This article presents a thorough treatment on the statistical analysis based on the number of concordant relative pairs in testing familial aggregation of clinical characteristics. The statistic has been used in the gastroenterology literature, with an asymptotic reference distribution constructed from population based estimates. The validity of the asymptotic normal approximation is a concern for traits with prevalence p close to zero. In the case of dichotomous trait, by comparing the asymptotic and exact distribution of the test statistic, we give guidelines for when the normal approximation is appropriate. For example, for n families of size 3, we need $np^2 > 5$. That implies, if $n = 100$, a good normal approximation demands a trait prevalence $p > 0.224$. The permutation method however, is not affected by rare trait prevalence. It is thus recommended for its fewer assumptions, robustness, and ease of implementation.

We have focused on the analysis of concordance counts with which gastroenterological researchers are familiar, but it is worth pointing out other statistical methods applicable for familial aggregation in such a context. Data from familial studies are clustered data with each family being a cluster. When within-cluster association is of explicit interest, the generalized estimating equations (GEE) approach can model the association parameter in addition to the mean structure (Yan and Fine, 2004). The GEE approach is robust in that only the mean and the covariance, instead of the full distribution, of the multivariate variable within cluster are specified. Unlike the permutation analysis of concordance counts where covariate effects can only be controlled in a limited way, GEE models allow covariates into both the mean and the association parameters through appropriate link functions. For illustration, consider the dichotomous trait terminal ileum involvement (termileum) for CD patients. We can use a logistic regression to model the probability of having terminal ileum involvement with covariates gender, age, and disease duration. To model the within family association, we specify the pairwise log odds ratio by a regression model with indicators of relative pair types as covariates. With the R package geepack (Halekoh et al., 2006), the estimated log odds ratio (standard error) are 1.556(0.356) for sib pairs, 1.045(0.796) for parent-child pairs, -0.483(2.247) for grandparent-grandchild pairs, 1.943(0.904) for uncle/aunt-nephew/niece pairs, and -0.945(3.565) for spousal pairs. These results are consistent with those in Table 3:

the log odds ratio is significantly nonzero between sib pairs but not so between parent-child pairs. Continuous and categorical traits can be modeled similarly. The conclusions about familial aggregation in all other clinical characteristics using the the GEE approach are comparable to those in Table 3, except in a few occasions where the GEE model estimates are unstable due to too few matched 1–1 pairs. GEE models have a clear advantage in incorporating covariate effects. Since the inference of GEE is based on asymptotic normality of the parameter estimate, one needs to be cautious when the trait is extremely rare and concordance is even rarer, in which case, a very large sample is needed for the asymptotics to be effective. Extreme rarity in traits and in concordance lead to unreliable parameter estimates since logit link function for prevalence and log link function for odds ratio explode near zero. Nevertheless, GEE can be applied to the IOIBD data and merits more attention in general from gastroenterologists.

Another class of models which may be useful is the generalized linear mixed model (GLMM) (e.g., Breslow and Clayton, 1993), also known as multilevel model (e.g., Goldstein, 1995). GLMM is a mixed effects extension of generalized linear models, with random effects added on the linear predictor. A random effect can be assumed at the family level which introduces association within family. In contrast to GEE, GLMM specifies the full likelihood. Normal distribution is usually assumed for random effects. Inference for GLMM is difficult because the likelihood involves integrating the random effects out, which cannot be computed explicitly in general. Widely used maximum likelihood methods are penalized quasi likelihood (PQL) (Breslow and Clayton, 1993) and Monte Carlo EM algorithm (McCulloch, 1997). The null hypothesis of no familial aggregation means that the variance of the random effect at the family level equals zero. Under this null hypothesis the variance parameter for the family level random effects is on the boundary of the parameter space, and, as a result, the likelihood ratio test statistic has a non-standard null distribution (Self and Liang, 1987). GLMM is able to accommodate association within clusters through random effects in estimating regression parameters in the mean model, but when associations are of explicit interest, GEE may be preferred for its milder assumption and easier implementation.

The permutation method for testing co-aggregation of two traits is admittedly a simple first thrust at the problem. For dichotomous traits, Hudson et al. (2001) proposed a multivariate logistic regression based on a quadratic exponential distribution. Within the framework of clustered data analysis, we can use GEE or GLMM for co-aggregation modeling as well. Keeping each family as a cluster, we can treat 2 or more traits from the same subject as a sub-cluster. A working covariance structure models the association between the traits can be assumed to carry out the GEE approach. Multivariate random effects at the subject level, in addition to the random effects at the family level, can be assumed to carry out the GLMM approach. Needless to say, implementation the likelihood inference of such a GLMM approach with multiple random effects can be challenging. A compromise is to use the composite likelihood (Lindsay, 1988), where, for example, pairwise bivariate distributions are composed together without fully specifying the joint distribution within cluster. Co-aggregation analysis with composite likelihood is an interesting direction for future research.

In the IOIBD study, with data collected from multi-centers world wide, cultural differences and genetic heterogeneity may be considered after the familial aggregation in various traits has been detected. King et al. (1984) approach genetic epidemiological studies from

the point of view of three sequential questions: 1) Do diseases cluster in families? 2) Is familial clustering related to common environmental exposure, biologically inherited susceptibility, or cultural inheritance of risk factors? 3) How is genetic susceptibility inherited? The classical approach to the problem has been to move sequentially through these questions with varying study designs and analytical methods. If familial aggregation is demonstrated, it would be appropriate to then explore the secondary questions. The IOIBD Multiplex Family Study in its first completed phase reported here has dealt exclusively with the first question, with respect to trait clustering within IBD families. Familial aggregation is a proxy of shared genetic and/or environmental influences upon those traits. Thus, by design our study's demonstration of aggregation cannot decipher the relative effects of these two influences. Some limited inferences regarding this are possible by comparing generational subset concordances (i.e. sib-sib pair concordances versus parent-child concordances). More importantly however, the demonstration that some traits are familial and others are not, can properly direct future studies aimed more specifically at analyzing the relative importance of genes, environment, and cultural effects upon disease phenotypes. A starting point would be the first consideration of those traits most strongly shared among families, with either exclusion of non-familial traits or stratification/controlling for non-shared traits which might act as confounders or contribute to unwanted heterogeneity in genetic association studies.

Our study has intentionally not attempted to address the problem of common cultural effects which might be important, considering that the data have been collected from 17 centers world-wide. The deciphering of genes versus environment in family-based study designs can be addressed either by study design or analytical methods. In design, the simplest method to infer cultural influences would be twin studies. However, twins are usually raised in the same environment, and while separated twins may address this issue of cultural effects, there are obvious practical difficulties for rare traits or diseases in identifying, recruiting and measuring exposures in sufficient numbers of affected separated twins. Analytic approaches to the problem require the collection of well-standardized environmental data, which is more properly addressed in a separate study than the familial aggregation study described here. Indeed, Khoury, Beaty, and Cohen haved stated that "Statistical modeling on family data without regard to such environmental factors cannot satisfactorily consider, adjust for, or rule out the presence of specific environmental agents affecting risk unless such information is systematically collected and examined for relatives as well as probands" (Khoury et al., 1993, p.192). Although in theory this could have been possible in the first phase of the IOIBD study, there would be significant resource limitations. Furthermore, the median duration of disease in this cohort is long (16.9 years) and thus it is unlikely that retrospective environmental data collection would be either practical or accurate.

In epidemiology, genetic heterogeneity refers to the phenomenon whereby multiple gene loci or mutations may lead to the same phenotype. Based upon results of many genetic linkage studies, association studies and genome-wide scans in different populations world-wide, this phenomenon is well recognized as regards IBD susceptibility, especially CD (Rioux et al., 2007; Negoro et al., 2003). Certainly, genetic heterogeneity is present in this cohort of multicenter subjects but without genotyping analysis, it is impossible to determine to what extent this has occurred. Nonetheless, our demonstration of familial concordance in traits remains valid, regardless of the degree of genetic heterogeneity. Population stratification is a similar yet distinct situation, where ethnicity may be considered as a confounder in genetic associ-

ation studies. In this case, a particular genotype under study may appear to be associated with the outcome, but the association may be confounded by the genotype's association with a particular ethnicity. This issue is possibly a concern in the secondary analyses of familial aggregation, where particular candidate genes are considered. The permutation method of concordance measurement described here cannot distinguish true risk factors from such confounders, but it has been convincingly argued that population stratification may not be of major importance when the number of ethnic groups in a study population is high, and that several alternative explanations are likely to account for failure to demonstrate significance in those studies where population stratification has been proposed as a problem (Wacholder et al., 2000). Obviously, further analyses of this IBD cohort should consider the issue further. A first step might involve matching or adjustment of cases and controls by ethnicity across the multiple study centers. However, if a true ethnic confounding exists, only adjusting for ethnicity is unlikely to satisfactorily eliminate bias due to population stratification. For further discussion of these issues, the reader is referred to Wacholder et al. (2002).

The proposed permutation method has been successfully applied to the IOIBD Multiplex Families Project. It is very flexible and has been adapted to solve practical issues such as non-dichotomous traits, special subgroups of relatives, and association of concordance in trait 1 and in trait 2. Although the permutation method is limited in its capability to incorporate covariates, it requires fewer assumptions and so can provide analysis complementary to the model based methods. The method is easy to implement in various computing environments. We hope to see more refinements, variations, and applications of this method in testing concordance of clinical traits in familial studies.

APPENDIX

A. Derivation of the mean and variance of N_k

Recall that the family size is k and the prevalence of the clinical characteristic is p. The number of concordance pairs N_k depends on W, the number of members with $Z = 1$. The distribution of W is binomial with parameter k and p. For $W < 2$, $N_k = 0$. For $W = i \geq 2$, $N_k = \binom{i}{2}$. Therefore,

$$E(N_k) = \sum_{i=2}^{k} \binom{i}{2} \binom{k}{i} p^i (1-p)^{(k-i)}.$$

Simplification gives $E(N_k) = \binom{k}{2} p^2$. The variance of N_k is obtained by getting $E(N_k^2)$ first,

$$E(N_k^2) = \sum_{i=2}^{k} \binom{i}{2}^2 \binom{k}{i} p^i (1-p)^{(k-i)}.$$

Simplification and $E(N_k)$ gives (2.1).

B. Sample Mathematica Code

The exact method can be implemented using software with symbolic calculation. A sample Mathematica session is presented in Figure 3. A function `myPGF` is defined to return the PGF

```
myPGF[p_, k_, n_] :=
 ((1 - p)^k + k (1 - p)^(-1+k) p + Sum[(1 - p)^(-i+k) p^i Z^(1/2 (-1+i) i) Binomial[k, i], {i, 2, k}])^n

CoefficientList[myPGF[0.2, 3, 20] myPGF[0.2, 4, 5], Z]
```

```
{0.0410308, 0.126389, 0.186346, 0.188784, 0.157975, 0.116801, 0.0774566, 0.0474146,
 0.027437, 0.0150296, 0.00784785, 0.00395114, 0.00191655, 0.000897335, 0.000408655,
 0.000180867, 0.0000778016, 0.0000327143, 0.0000134341, 5.38326×10^-6, 2.11494×10^-6,
 8.13887×10^-7, 3.064×10^-7, 1.1331×10^-7, 4.11233×10^-8, 1.46236×10^-8, 5.11474×10^-9,
 1.75774×10^-9, 5.92444×10^-10, 1.9656×10^-10, 6.4123×10^-11, 2.05265×10^-11, 6.47109×10^-12,
 2.00661×10^-12, 6.10655×10^-13, 1.83057×10^-13, 5.39786×10^-14, 1.56185×10^-14,
 4.4517×10^-15, 1.24787×10^-15, 3.43104×10^-16, 9.29172×10^-17, 2.47365×10^-17,
 6.45548×10^-18, 1.65898×10^-18, 4.18847×10^-19, 1.03581×10^-19, 2.52193×10^-20,
 6.02753×10^-21, 1.40983×10^-21, 3.24618×10^-22, 7.33×10^-23, 1.61821×10^-23, 3.51703×10^-24,
 7.48709×10^-25, 1.55669×10^-25, 3.18762×10^-26, 6.38282×10^-27, 1.24698×10^-27,
 2.40117×10^-28, 4.51062×10^-29, 8.25913×10^-30, 1.49254×10^-30, 2.62164×10^-31,
 4.48611×10^-32, 7.59107×10^-33, 1.24117×10^-33, 1.97835×10^-34, 3.12587×10^-35,
 4.72649×10^-36, 6.99338×10^-37, 1.02783×10^-37, 1.42295×10^-38, 1.94922×10^-39,
 2.64761×10^-40, 3.30391×10^-41, 4.19363×10^-42, 5.19312×10^-43, 5.70255×10^-44,
 6.77979×10^-45, 7.39541×10^-46, 6.9216×10^-47, 7.96034×10^-48, 6.9467×10^-49, 5.48562×10^-50,
 6.41214×10^-51, 3.30762×10^-52, 2.80471×10^-53, 2.90142×10^-54, 0, 1.20893×10^-56}
```

```
Export["p-0.4-n-200.csv", CoefficientList[myPGF[0.4, 3, 200], Z], "Table"]
```

```
p-0.4-n-200.csv
```

Figure 3: A sample Mathematica session to obtain the exact distribution of the number of concordant relative pairs in 20 families of size 3 and 5 families of size 4.

of X_{n_k} for n families of size k with clinical character prevalence p. For example, assuming $p = 0.2$, `myPGF(0.2, 3, 20)` gives the PGF of the number of concordant relative pairs in 20 families of size 3; `myPGF(0.2, 4, 5)` gives the PGF of the number of concordant relative pairs in 5 families of size 4. Taking `CoefficientList` of product of these two terms gives a table of the probabilities that the number of concordant relative pairs X in 20 families of size 3 and 5 families of size 4 takes value $x = 0, 1, 2, \ldots$. These probabilities are then dumped into an external file by `Export`.

REFERENCES

Bayless, T. M., Tokayer, A. Z., Polito, J. M., Quaskey, S. A., Mellits, E. D. and Harris, M. L. (1996). Crohn's disease: concordance for site and clinical type in affected family members–potential hereditary influences. *Gastroenterology* **111**, 573–579.

Berger, V. W. (2000). Pros and cons of permutation tests in clinical trials. *Statistics in Medicine* **19**, 1319–1328.

Besag, J. and Clifford, P. (1991). Sequential Monte Carlo p-values. *Biometrika* **78**, 301–304.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.

Goldstein, H. (1995). *Multilevel Statistical Models*. Edward Arnold Publishers Ltd.

Good, P. I. (2001). *Resampling Methods: a Practical Guide to Data Analysis*. Birkhauser Verlag.

Halekoh, U., Højsgaard, S. and Yan, J. (2006). The R package geepack for generalized estimation equation modeling. *Journal of Statistical Software* **15**, 1–11.

Holland, P. W. (1986). Statistics and causal inference (C/R: P961-970). *Journal of the American Statistical Association* **81**, 945–960.

Hudson, J. I., Laird, N. M. and Betensky, R. A. (2001). Multivariate logistic regression for familial aggregation of two disorders. I. development of models and methods. *American Journal of Epidemiology* **153**, 500–505.

Hugot, J. P., Cezard, J. P., Colombel, J. F., Belaiche, J., Almer, S., Tysk, C., Montague, S., Gassull, M., Christensen, S., Finkel, Y., Gower-Rousseau, C., Modigliani, R., Zouali, H., Lesage, S., Chamaillard, M., Macry, J., Thomas, G. and Victor, J. M. (2002). Clustering of Crohn's disease within affected sibships. *European Journal of Human Genetics* **11**, 179–184.

Hunt, S. C., Hasstedt, S. J., Williams, R. R. and Rao, D. C. (1986). Testing for familial aggregation of a dichotomous trait. *Genetic Epidemiology* **3**, 299–312.

Khoury, M. J., Beaty, T. H. and Cohen, B. H. (1993). *Fundamentals of Genetic Epidemiology*. Oxford University Press.

King, M. C., Lee, G. M., Spinner, N. B., Thomson, G. and Wrensch, M. R. (1984). Genetic epidemiology. *Annual Review of Public Health* **5**, 1–52.

Lee, J. C. W. and Lennard-Jones, J. E. (1996). Inflammatory bowel disease in 67 families each with three or more affected first-degree relatives. *Gastroenterology* **111**, 587–596.

Lindsay, B. G. (1988). Composite likelihood methods. In N. U. Prabhu (ed.), *Statistical Inference from Stochastic Processes*, pp. 221–239. American Mathematical Society.

Little, R. J. and Rubin, D. B. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. *Annual Review of Public Health* **21**, 121–145.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. John Wiley & Sons.

Louis, E., Michel, V., Hugot, J. P., Reenaers, C., Fontaine, F., Delforge, M., El Yafi, F., Colombel, J. F. and Belaiche, J. (2003). Early development of stricturing or penetrating pattern in Crohn's disease is influenced by disease location, number of flares, and smoking but not by NOD2/CARD15 genotype. *Gut* **52**, 552–557.

Lynch, H. T., Fain, P. R., Golgar, D., Albano, W. A., Mailliard, J. A. and McKenna, P. (1981). Familial breast cancer and its recognition in an oncology clinic. *Cancer* **47**, 2730–2739.

Lynch, H. T., Kimberling, W. J., Biscone, K. A., Lynch, J. F., Wagner, C. A., Brennan, K., Mailliard, J. A., Johnson, P. S., Soori, J. S. and McKenna, P. J. (1986). Familial heterogeneity of colon cancer risk. *Cancer* **57**, 2089–2096.

Maldonado, G. and Greenland, S. (2002). Estimating causal effects (C/R: P429-438). *International Journal of Epidemiology* **21**, 422–429.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* **92**, 162–170.

Negoro, K., McGovern, D. P., Kinouchi, Y., Takahashi, S., Lench, N. J., Shimosegawa, T., Carey, A., Cardon, L. R., Jewell, D. P. and van Heel, D. A. (2003). Analysis of the IBD5 locus and potential gene-gene interactions in crohn's disease. *Gut* **52**, 541–546.

Nettleton, D. and Doerge, R. W. (2000). Accounting for variability in the use of permutation testing to detect quantitative trait loci. *Biometrics* **56**, 52–58.

Peeters, M., Nevens, H., Baert, F., Hiele, M., de Meyer, R., AMand Vlietinck and Rutgeerts, P. (1996). Familial aggregation in Crohn's disease: increased age-adjusted risk and concordance in clinical characteristics. *Gastroenterology* **111**, 597–603.

Pesarin, F. (2001). *Multivariate Permutation Tests: with Applications in Biostatistics*. John Wiley & Sons.

Picco, M., Goodman, S., Reed, J. and Bayless, T. M. (2001). Methodological pitfalls in the determination of genetic anticipation: The case of Crohn disease. *Annals of Internal Medicine* **134**, 1124–1129.

Rioux, J. D., Xavier, R. J., Taylor, K. D., Silverberg, M. S., Goyette, P., Huett, A., Green, T., Kuballa, P., Barmada, M. M., Datta, L. W., Shugart, Y. Y., Griffiths, A. M., Targan, S. R., Ippoliti, A. F., Bernard, E. J., Mei, L., Nicolae, D. L., Regueiro, M., Schumm, L. P., Steinhart, A. H., Rotter, J. I., Duerr, R. H., Cho, J. H., Daly, M. J. and Brant, S. R. (2007). Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nature Genetics* **39**, 596–604.

Rubin, D. B. (2006). *Matched Sampling for Causal Effects*. New York: Cambridge University Press.

Russell, R. K. and Satsangi, J. (2004). IBD: A family affair. *Best Practice & Research Clinical Gastroenterology* **18**, 525–539.

Schwartz, A. G., Boehnke, M. and Moll, P. P. (1988). Family risk index as a measure of familial heterogeneity of cancer risk. A population-based study in metropolitan Detroit. *American Journal of Epidemiology* **128**, 524–535.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605–610.

Tamboli, C. P., Colombel, J. F., Hillis, S. L., Attaluri, P. A. and Duke, A. (2005). Phenotypic concordance and time clustering of diagnosis in multiply affected IBd families: The IOIBD Multiplex Family Collaborative Project, Phase I & II. *Gastroenterology* **128**, A327.

Wacholder, S., Rothman, N. and Caporaso, N. (2000). Population stratification in epidemiologic studies of common genetic variants and cancer: Quantification of bias. *Journal of National Cancer Institute* **92**, 1151–1158.

Wacholder, S., Rothman, N. and Caporaso, N. (2002). Counterpoint: Bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. *Cancer Epidemiol Biomarkers Prevention* **11**, 513–520.

Wolfram Research, Inc. (2005). *Mathematica (Version 5.2)*. Wolfram Research, Inc., Champaign, Illinois.

Yan, J. and Fine, J. P. (2004). Estimating equations for association structures. *Statistics in Medicine* **23**, 859–874.